

## CHARACTERIZING THE ERROR IN THE ESTIMATED AGE-DEPTH RELATIONSHIP

Andrew R Solow

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. Email: asolow@whoi.edu.

**ABSTRACT.** It is common practice to estimate the age of undated material extracted from a sediment core from radiocarbon or other radiometric dates of samples taken above and below the extracted material. This paper presents a simple expression for the variance of this estimated age. This variance accounts for both  $^{14}\text{C}$  dating error and error due to bioturbation.

### INTRODUCTION

It is common practice to estimate the continuous age-depth relationship along a sediment core from radiocarbon or other radiometric dates of samples taken at discrete depths in the core. Such a relationship can then be used to estimate the age of undated material extracted at a known depth in the core. In doing so, it may be useful to have an idea of the magnitude of the error in the estimated age of the extracted material. For example, this would be important in assessing the significance of differences in the estimated timing of climatic events reconstructed by a geochemical analysis of material extracted from different cores. The purpose of this paper is to describe a simple model of this estimation error. The model, which appears to be the first of its kind, includes both  $^{14}\text{C}$  dating error and error due to vertical sediment mixing.

The remainder of the paper is organized in the following way. The basic model is developed in the next section. In the third section, this model is applied to a small data set. The final section contains some concluding remarks.

### AN ERROR MODEL

The basic problem considered here is the estimation of the calendar age  $Y_0$  of material extracted at nominal depth  $d_0$  within a core. This paper focuses on the simple, but common, case where estimation is by linear interpolation between 2 control points at nominal depths  $d_1$  and  $d_2$  with  $d_1 \leq d_0 \leq d_2$ . In this case, the estimate of  $Y_0$  is:

$$\hat{Y}_0 = w\hat{Y}_1 + (1 - w)\hat{Y}_2 \quad (1)$$

where  $\hat{Y}_1$  and  $\hat{Y}_2$  are the estimated calendar ages of material extracted at the control points and:

$$w = (d_2 - d_0) / (d_2 - d_1) \quad (2)$$

Linear interpolation is based on the assumption that, between control points, calendar age increases approximately linearly with depth.

The goal of this paper is to provide an expression for the variance of the estimation error  $Y_0 - \hat{Y}_0$ . Two sources of error contribute to this variance. First, as a result of bioturbation or other reworking of sediment, the true depositional depth of material extracted at the control and estimation points may differ from their nominal depths. This will be referred to as depth error. Second, the estimated calendar ages of the material extracted at the control points are subject to the usual dating errors. To proceed further, it is necessary to specify a statistical model of these errors and the way in which they interact.

Let the random variable  $Z(d)$  be the true depth of material collected at nominal depth  $d$ . Assume that:

$$Z(d) = d + \varepsilon \quad (3)$$

where  $\varepsilon$  is a depth error with mean 0 and variance  $\sigma_\varepsilon^2$ . A model of this variance is given below. Suppose that, as at the control points, the calendar age of this material is estimated by calibrating an estimate of its  $^{14}\text{C}$  age. Let:

$$\hat{R} = r[Z(d)] + \eta \quad (4)$$

be the estimated  $^{14}\text{C}$  age of this material where, in general,  $r(z)$  is the true  $^{14}\text{C}$  age of material at true depth  $z$ , and  $\eta$  is the  $^{14}\text{C}$  dating error with mean 0 and known variance  $\sigma_\eta^2$ . Provided  $\varepsilon$  is small in relation to the scale of curvature of  $r(z)$ :

$$\begin{aligned} \hat{R} &\cong r(d) + r'(d)\varepsilon + \eta \\ &= r(d) + e \end{aligned} \quad (5)$$

where  $r'$  is the derivative of  $r(z)$  with respect to  $z$ . This derivative is essentially the reciprocal of the sedimentation rate. The combined error  $e$  has mean 0 and variance:

$$\sigma_e^2 = r'(d)^2 \sigma_\varepsilon^2 + \sigma_\eta^2 \quad (6)$$

The estimated calendar age is given by:

$$\hat{Y} = c(\hat{R}) \quad (7)$$

where  $c(r)$  is a fixed calibration curve. Provided that  $e$  is small in relation to the curvature of  $c(r)$ :

$$\begin{aligned} \hat{Y} &\cong c[r(d)] + c'[r(d)]e \\ &= c[r(d)] + f \end{aligned} \quad (8)$$

where  $c'$  is the derivative of  $c(r)$  with respect to  $r$  and  $f$  is the calendar age error with mean 0 and variance:

$$\sigma_f^2 = c'[r(d)]^2 [r'(d)^2 \sigma_\varepsilon^2 + \sigma_\eta^2] \quad (9)$$

The quantity  $c[r(d)]$  in (8) is the calibrated value of the true  $^{14}\text{C}$  age of material, of which the true depositional depth is equal to the nominal depth  $d$ . Thus, under this model, the estimated calendar age  $\hat{Y}$  is unbiased with variance given in (9). The mean of the true calendar age  $Y$  of material extracted at nominal depth  $d$  is also  $c[r(d)]$ , but its variance is:

$$\sigma_Y^2 = c'[r(d)]^2 r'(d)^2 \sigma_\varepsilon^2 \quad (10)$$

The difference between (9) and (10) is due to the fact that  $Y$  (which is unobservable) involves no dating and, therefore, no dating error, while  $\hat{Y}$  (which is observable) does.

Returning to the main goal of this paper, it follows from these results that  $\hat{Y}_\theta$  is an unbiased estimate of  $Y_\theta$ . Under the assumption that the depth errors at the control and estimation points, and the dating errors at the control points, are independent, the variance of the estimation error  $Y_\theta - \hat{Y}_\theta$  is:

$$\text{Var}(Y_\theta - \hat{Y}_\theta) = \text{Var}(Y_\theta) + w^2 \text{Var}(\hat{Y}_1) + (1 - w)^2 \text{Var}(\hat{Y}_2) \quad (11)$$

Finally, it follows from (9) and (10) that this variance is given by:

$$\begin{aligned} \text{Var}(Y_0 - \hat{Y}_0) = & c'[r(d_0)]^2 r'(d_0)^2 \sigma_{\varepsilon_0}^2 + w^2 c'[r(d_1)]^2 [r'(d_1)^2 \sigma_{\varepsilon_1}^2 + \sigma_{\eta_1}^2] + \\ & (1-w)^2 c'[r(d_2)]^2 [r'(d_2)^2 \sigma_{\varepsilon_2}^2 + \sigma_{\eta_2}^2] \end{aligned} \quad (12)$$

where  $\sigma_{\varepsilon_0}^2$ ,  $\sigma_{\varepsilon_1}^2$ , and  $\sigma_{\varepsilon_2}^2$  are the variances of the depth errors at the estimation and control points and  $\sigma_{\eta_1}^2$  and  $\sigma_{\eta_2}^2$  are the variances of the  $^{14}\text{C}$  dates at the control points.

Before turning to a simple model of depth error, it is worth noting that the estimation variance in (11) depends through  $w$  on the relative location of the estimation point between the 2 control points. For example, if  $\text{Var}(\hat{Y}_1) = \text{Var}(\hat{Y}_2)$ , then  $\text{Var}(Y_0 - \hat{Y}_0)$  is maximized when  $d_0 = d_1$  or  $d_0 = d_2$  (so that  $w = 1$  or  $0$ ), and minimized when  $d_0 = (d_1 + d_2) / 2$  (so that  $w = 1/2$ ). Briefly, estimating  $Y_0$  by averaging  $\hat{Y}_1$  and  $\hat{Y}_2$  reduces variance by averaging the errors in these estimated ages. When the estimation point is close to one of the control points, the estimate of  $Y_0$  is dominated by the estimated age at that control point with little averaging of errors. In contrast, when the estimation point is midway between the control points, the estimated ages at the control points receive equal weight in  $\hat{Y}_0$  with maximal averaging of errors.

The estimation error variance in (12) depends on the variance of the depth errors at the control and estimation points. A simple model of these depth errors is the following. The basic assumption is that the material collected at nominal depth  $d$  consists of  $n$  particles distributed at random within a sample volume of uniform height  $h$  centered at  $d$ . As a result of sediment reworking, between the time of deposition and the time of collection, each particle has undergone a random vertical movement of magnitude  $\Delta$  with mean 0 and variance  $\sigma_{\Delta}^2$ . The variance of the true depth of a single such particle is the sum of the variance  $h^2 / 12$  due to its random location within the sample volume and the variance  $\sigma_{\Delta}^2$  due to sediment reworking. The true depth  $Z(d)$  of the sample material can be approximated by the average of the true depths of  $n$  of these particles. This average has mean  $d$  and variance:

$$\sigma_{\varepsilon}^2 = [\sigma_{\Delta}^2 + (h^2 / 12)] / n \quad (13)$$

In many cases,  $h \ll \sigma_{\Delta}$ , so that  $\sigma_{\varepsilon}^2 \cong \sigma_{\Delta}^2 / n$ .

## AN APPLICATION

Table 1 reports the depth,  $^{14}\text{C}$  age and error ( $\sigma_{\eta}$ ), and calibrated calendar age of 12 samples of the planktonic foraminifer *Globigerinoides sacculifer* in a core taken near the Bahamas at 26°04'N, 78°03'W, in approximately 1000 m of water. These data and subsidiary information about them were kindly provided by William Curry. The continuous age-depth relationship for this core estimated by linear interpolation is shown in Figure 1. For these data, the calibration curve is essentially linear, with slope  $c'(r) = 1.128$  for all values of  $r$ . The  $^{14}\text{C}$  age at each control point was based on  $n \cong 100$  specimens, extracted from a sediment wedge of height  $h = 0.02$  m. The bioturbation error  $\sigma_{\Delta}$  is around 0.05 m. Thus, from (13), the standard deviation  $\sigma_{\varepsilon}$  of the depth error at each control point is approximately 0.005 m.

To illustrate the calculation of the variance in (12), consider estimating the age of a single particle extracted from a sediment wedge of height 0.02 m centered at depth  $d_0 = 2.1$  m, based on the estimated calendar ages at the control points at  $d_1 = 2.0$  m and  $d_2 = 2.2$  m. It follows from (13) that the standard deviation  $\sigma_{\varepsilon_0}$  of the depth error at the estimation point is around 0.05 m. That this is an order of magnitude larger than  $\sigma_{\varepsilon}$  at each control point reflects the effect of averaging 100 particles at each control point. From (1) and (2), the estimated age  $\hat{Y}_0$  is 38.7 kyr. To complete the calculation

Table 1 Depth, radiometric age and error, and calibrated calendar age for 12 samples of *G. sacculifer* picked from a sediment core.

Depth (m)	Radiometric age $\pm \sigma_\eta$ (yr)	Calendar age (yr)
0.10	920 $\pm$ 35	510
0.62	5290 $\pm$ 45	5640
0.88	7630 $\pm$ 45	7930
1.13	11,000 $\pm$ 50	12,500
1.21	12,200 $\pm$ 55	13,800
1.34	17,100 $\pm$ 100	19,700
1.51	20,200 $\pm$ 85	23,400
1.71	25,900 $\pm$ 120	29,800
2.00	31,500 $\pm$ 170	35,800
2.20	37,400 $\pm$ 360	41,600
2.40	39,600 $\pm$ 390	43,300
2.67	45,700 $\pm$ 500	49,300

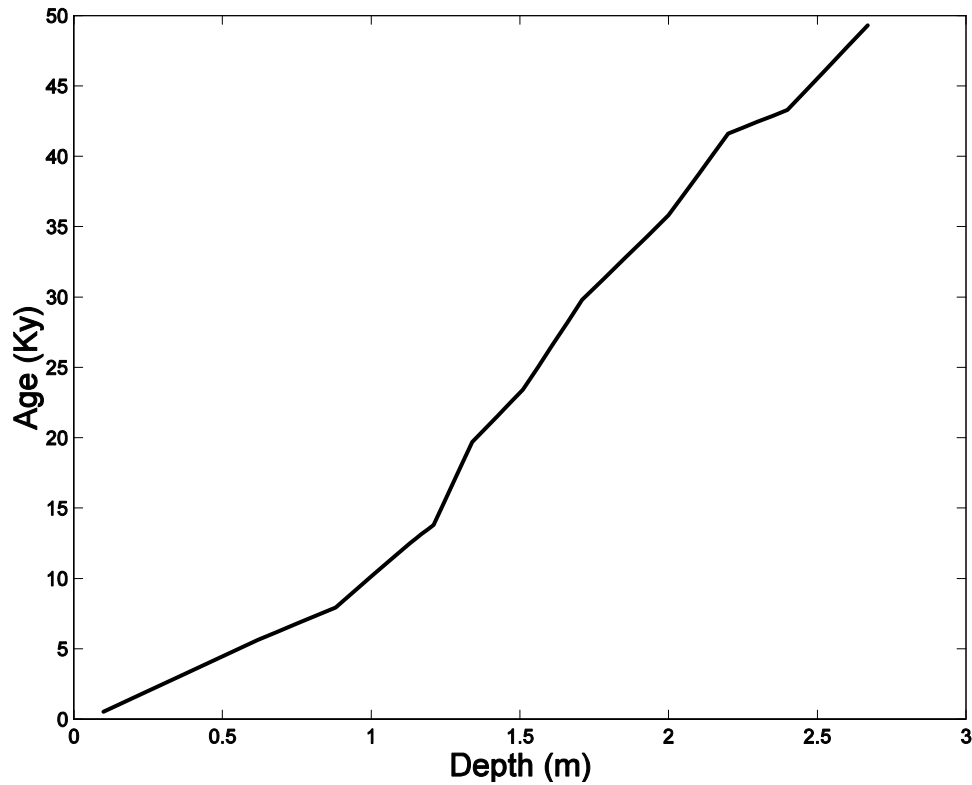


Figure 1 Age-depth curve estimated by simple linear interpolation between control points.

in (12), it is necessary to estimate  $r'(d)$  at  $d_0$ ,  $d_1$ , and  $d_2$ . Under the assumption that the sedimentation rate is approximately constant between control points, a simple estimate is  $[r(d_2) - r(d_1)] / (d_2 - d_1)$  or 29.5 kyr  $m^{-1}$ . Combining terms according to (12) yields  $Var(Y_0 - \hat{Y}_0)^{1/2} = 1.69$  kyr.

In Figure 2,  $\text{Var}(Y_0 - \hat{Y}_0)^{1/2}$  is plotted against  $d_0$  the entire core. Although the standard deviation  $\sigma_n$  of the  $^{14}\text{C}$  dating error varies by a factor of nearly 15, the dominant source of variability in this figure is variability in  $r'(d)$ . The reason is that the low sedimentation in relation to the level of bioturbation, combined with the absence of particle averaging at the estimation point, make  $c'[r(d_0)] r'(d_0) \sigma_{\varepsilon_0}$  the dominant contributor to estimation error. The only part of this term that varies over the core is  $r'(d_0)$ .

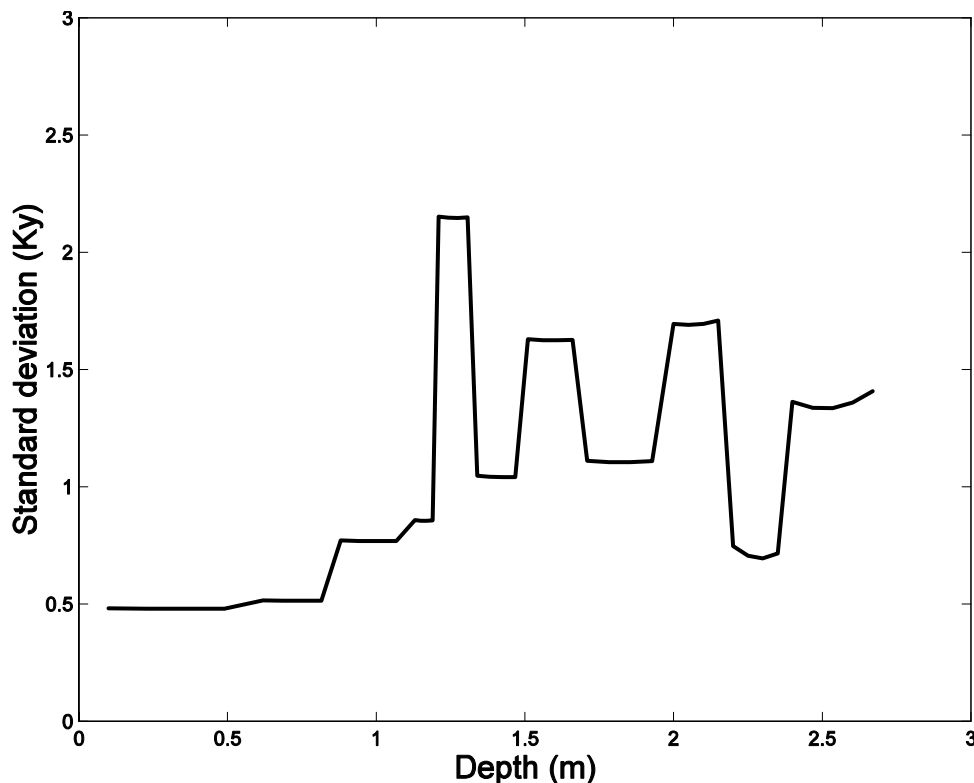


Figure 2 Approximate standard deviation of estimation errors.

## DISCUSSION

The purpose of this paper has been to describe and illustrate a way to characterize error in the age-depth relationship. This characterization appears to be the first of its kind. It is based on 2 assumptions of approximate local linearity: specifically, that the  $^{14}\text{C}$  age-depth relationship  $r(z)$  is approximately locally linear in depth  $z$  at the scale of the depth error  $\varepsilon$  and that the calibration curve  $c(r)$  is approximately locally linear in  $^{14}\text{C}$  age  $r$  at the scale of the total  $^{14}\text{C}$  age error  $e$ . Another key assumption is that, at each control point, the deviations of the true depths of the particles contained within the sampling volume from their common nominal depth are independent. This assumption would be violated if this material had been subjected to a large-scale disturbance that affected all of the particles in the same way. In a specific application, the reasonableness of these assumptions should be established before proceeding.

Beyond its use as a quantitative measure of estimation error, the error characterization presented here can provide useful qualitative information about the sources of this error. For example, in the application described in the previous section, it is clear that improving the  $^{14}\text{C}$  dating at the control points would have little benefit in estimating the age of undated material.

Finally, it is possible to extend the error model described in this paper in at least 2 directions. First, this paper has focused on the simple (but common) case in which estimation is based on linear interpolation between 2 control points. It is straightforward to apply the error model to any other estimator that is a linear function of the calibrated ages of the control points. Such estimators include those based on regression and splines. Some details are provided in an appendix. Second, the model treats the calibration curve as fixed. It would also be possible to extend the model to include the effect of error in the calibration curve.

## ACKNOWLEDGEMENTS

William Curry provided the data used in this paper. The helpful comments of an anonymous reviewer are acknowledged with gratitude.

## APPENDIX

This appendix gives an expression for the error variance for a general linear estimate of the age-depth relationship. Let  $\hat{Y} = (\hat{Y}_1 \hat{Y}_2 \dots \hat{Y}_n)^t$  be the vector of estimated calendar ages of material extracted at  $n$  control points. Here and below, the superscript  $t$  denotes the vector or matrix transpose. A general linear estimate of the calendar age  $Y_0$  of undated material extracted at nominal depth  $d_0$  has the form:

$$\hat{Y}_0 = w^t \hat{Y} \quad (\text{A.1})$$

where  $w = (w_1 w_2 \dots w_n)^t$  is a vector of weights. The variance of the error  $Y_0 - \hat{Y}_0$  is:

$$\text{Var}(Y_0 - \hat{Y}_0) = \text{Var}(Y_0) + \text{Var}(\hat{Y}_0) \quad (\text{A.2})$$

The first term on the righthand-side of (A.2) is given in Equation (10) in the paper. The second term is:

$$\text{Var}(\hat{Y}_0) = w^t \Sigma w \quad (\text{A.3})$$

where  $\Sigma$  is the  $n$ -by- $n$  diagonal matrix, with elements  $\text{Var}(Y_1), \text{Var}(Y_2), \dots, \text{Var}(Y_n)$  along the diagonal. These variances are given in Equation (9) in the paper.

As a simple example, suppose that the age-depth model is estimated by fitting the line:

$$Y_j = \beta_0 + \beta_1 d_j \quad (\text{A.4})$$

by ordinary least squares, where  $d_j$  is the depth of control point  $j$ . The ordinary least squares estimate of  $\beta = (\beta_0 \beta_1)^t$  is:

$$\hat{\beta} = (X^t X)^{-1} X^t Y \quad (\text{A.5})$$

where  $X$  is the  $n$ -by-2 matrix whose first column has elements all equal to 1 and second column has elements  $d_1, d_2, \dots, d_n$ . The estimate of  $Y_0$  is given by:

$$\hat{Y}_0 = x_0^t \hat{\beta} \quad (\text{A.6})$$

where  $x_0 = (1 \ d_0)^t$ . From (A.5), this estimate can be written in the form (A.1) with:

$$w = x_0^t (X^t X)^{-1} X^t \quad (\text{A.7}).$$